NLP Coursework Specs (2022)

Summary: Your task will be to develop a binary classification model to predict whether a text contains patronising or condescending language, writing a report based on your findings. The task is [task 4](#) (subtask 1) in SemEval 2022, with more information available from the [task paper](#).

Your task: Implement a transformer-based model from Huggingface that outperforms the task's [RoBERTa-base baseline](#) (provided by the organisers) in F1 score on the official dev dataset (0.48). We encourage creativity in your approach, so you can decide how to go about improving the current baseline performance. You can also reuse any existing code.

The coursework will be mostly marked based on your report, with some marks also awarded for your project code. While you are expected to improve on the performance of the [RoBERTa-base baseline](#), your report will be assessed on how well you answer the questions in the *Task Details* section below rather than on your raw model performance.

Deadline: The coursework deadline is the **4th March**

Data and evaluation: The training data can be downloaded from [this link](#). With an allocation of this data into train and official dev sets also provided [here](#). In addition, you should create an internal dev set (a portion of the training set) for hyperparameter tuning and model selection. As the test data is held out during the competition, creating an additional internal dev set allows you to use the official dev set to test performance. One member of your team must also submit [this form](#) prior to starting work on the coursework.

You will need to submit your final predictions on the test set to CodaLab, sharing your final model performance on the test set in your report alongside the performance of the model on the official dev set. To evaluate on the latter, you can either submit to the Practice section of CodaLab or evaluate offline.

Teams*:* You should form groups of up to 3 students. Submissions with fewer students are possible, but will *not* be assessed differently. Please choose a team leader to create an account on CodaLab. Specify on the profile the team name and team members (names of students). It is your responsibility to request access to join the competition with enough time for the organisers to add you - we recommend doing this at the beginning.

If you have issues with a team member not contributing you should inform us well in advance of the deadline (please email l.specia@imperial.ac.uk).

Code: You will be required to provide either a link to a GitHub repo or a Colab notebook in your report:
- Your code should run and train your model by running a single file without additional intervention, displaying your model performance (precision, recall and F1) on the official dev set.
- This should match the official dev set results provided in your report.

Report: Your report should be between 2 or 3 pages, using this template.

The 2022 NLP Coursework Prize

The top performing group will be awarded a prize for the best coursework submission. This will involve a physical prize, in addition to the opportunity to present the work to your peers. Two runner-up groups will also be awarded a prize.

The prize will be awarded based on the overall quality of the report and the depth of the analysis conducted.

Marking Scheme

1) Data analysis of the training data (15 marks):

For a written description of the training data. This should include:

1. **5 marks** – Analysis of the class labels: how frequent these are and how they correlate with any feature of the data, e.g. input length.
2. **10 marks** - Qualitative assessment of the dataset, considering either how hard or how subjective the task is, providing examples in your report.

2) Modelling (45 marks):

For the successful implementation of a HuggingFace model:

1. **10 marks**: Successful implementation of a HuggingFace transformer model (train and produce predictions which outperform the F1 score for the [RoBERTa-base baseline](#) provided - 0.48).
2. **10 marks**: Choices of data processing (e.g. deciding whether the model should be case-sensitive), how to use given training labels, and choice of model hyperparameters, e.g. an appropriate learning rate, learning schedule and if using early-stopping. You should justify the decisions that you make, including which model you're using.
3. **10 marks**: Creative direction for improvement (beyond using a bigger Transformer model), for example pre-processing, data sampling, data augmentation, ensembling, etc.
4. **5 marks**: Readability and "runnability" of your model code
   (marked from your project code)
5. **5 marks**: Correctness of this code compared to your report description
   (marked from your project code and report)
6. **5 marks**: Description of the model results and your hyper-parameter tuning (some evidence of this is required in your report), alongside the final model results for the task. Any figures provided in your report should clearly state if these are on your own internal dev set, the official dev set or the test set.

3) Analysis (15 marks):

Analysis questions to be answered (each will require running additional experiments):

Your report should state the analysis questions so that this can be read as a self-contained report, rather than referring to 'analysis question 1' etc.

1. ***5 marks***: To what extent is the model better at predicting examples with a higher level of patronising content? Justify your answer.

2. ***5 marks***: How does the length of the input sequence impact the model performance? If there is any difference, speculate why.
3. ***5 marks***: To what extent does the categorical data provided influence the model predictions?

4) Written report (25 marks):

Marks are awarded for the quality of your written report:

1. **5 marks**: Introduction, with an explanation of the task and the dataset. You may want to read/cite the task paper (and any other paper of your choosing).
2. **5 marks**: Readability of the report (language, coherence, etc.).
3. **10 marks**: Good use of graphs or results tables that address the analysis questions.
4. **5 marks**: Conclusion, with a summary of your results, and your key findings from the analysis questions. You may suggest further experiments as next steps.